GROUP 9 - AZIZ , LAMIA , MANAL

Statistics of Indian School Education

Presented by Group 9

PRESENTATION OVERVIEW

Concept
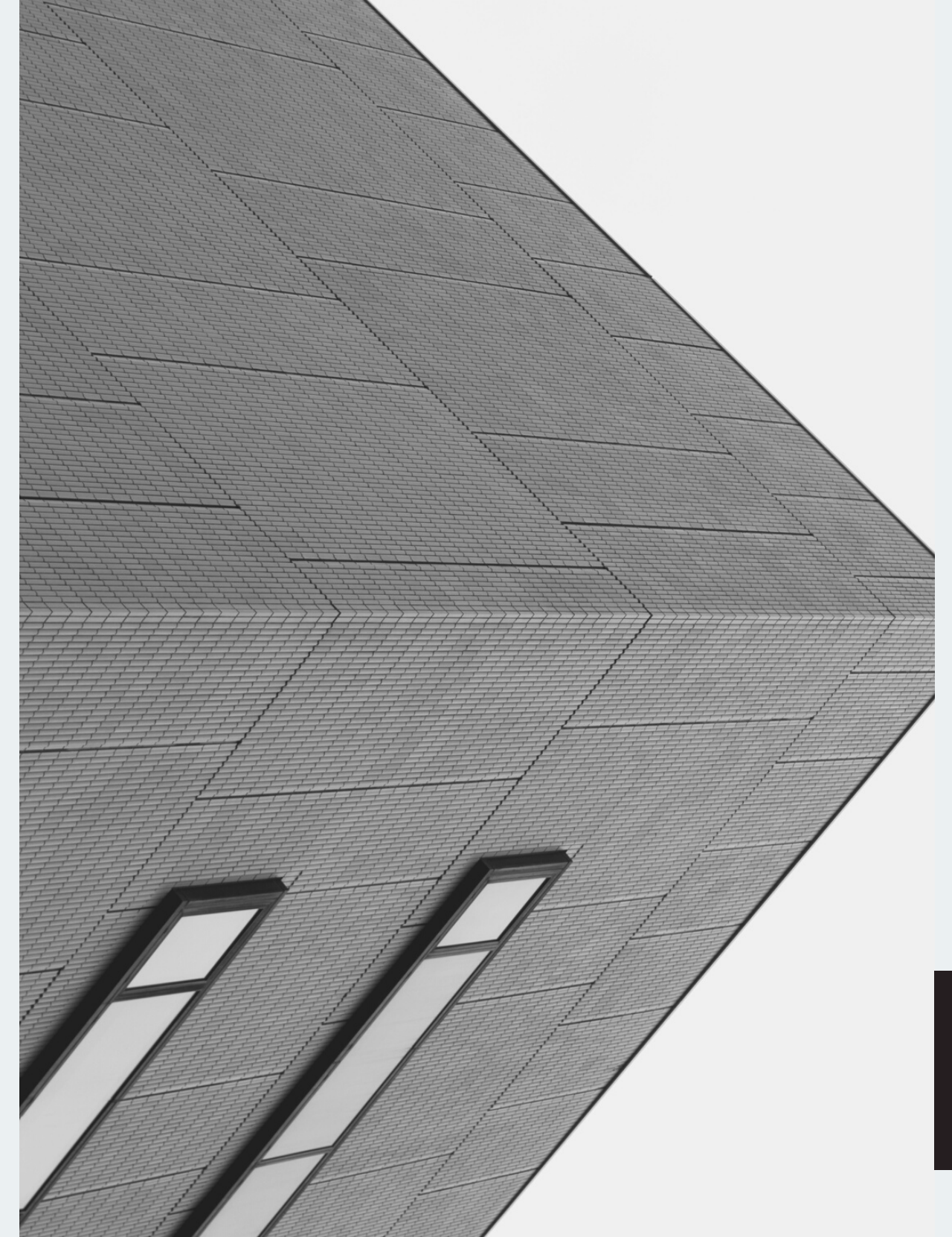Hypothesis
Facts
Figures
Prediction

LAMIA , MANAL , AZIZ

Group 9 .Sep 2020

CONCEPT

INDIAN SCHOOL QUANTITATIVE ANALYSIS

The objective goal is to burrow a few realities about the nature of Indian Education from 2013 - 2016 and help individuals of India and Government of India to see better about the Indian Schools to take quick activities if necessary.

Group 9 .Sep 2020

LEARNING IS A TREASURE

"Education is our passport to the future, for tomorrow belongs to the people who prepare for it today." –

MALCOLM X

HYPOTHESIS#1

STUDENT IN LESS DEVELOPED STATES DROP OUT MORE TO SUPPORT THEIR FAMILIES

```python
# Importing the Dataset

for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

data_dropout = pd.read_csv('../input/indian-school-education-statistics/dropout-ratio-2012-2015.csv')
enroll = pd.read_csv('../input/indian-school-education-statistics/gross-enrollment-ratio-2013-2016.csv')
data_comps = pd.read_csv('../input/indian-school-education-statistics/percentage-of-schools-with-comps-2013-2016.csv')
data_electricity = pd.read_csv('../input/indian-school-education-statistics/percentage-of-schools-with-electricity-2013-2016.csv')
data_water_facility = pd.read_csv('../input/indian-school-education-statistics/percentage-of-schools-with-water-facility-2013-2016.csv')
data_boys_toilet = pd.read_csv('../input/indian-school-education-statistics/schools-with-boys-toilet-2013-2016.csv')
data_girls_toilet = pd.read_csv('../input/indian-school-education-statistics/schools-with-girls-toilet-2013-2016.csv')
```

```
/kaggle/input/indian-school-education-statistics/percentage-of-schools-with-water-facility-2013-2016.csv
/kaggle/input/indian-school-education-statistics/percentage-of-schools-with-comps-2013-2016.csv
/kaggle/input/indian-school-education-statistics/schools-with-girls-toilet-2013-2016.csv
/kaggle/input/indian-school-education-statistics/percentage-of-schools-with-electricity-2013-2016.csv
/kaggle/input/indian-school-education-statistics/gross-enrollment-ratio-2013-2016.csv
/kaggle/input/indian-school-education-statistics/dropout-ratio-2012-2015.csv
/kaggle/input/indian-school-education-statistics/schools-with-boys-toilet-2013-2016.csv
```

+ Code   + Markdown

```python
# Create database
dbname = 'IndianEducationDb'
conn = sqlite3.connect(dbname + '.sqlite')

cur = conn.cursor()

# Load the data to SQL Database
data_dropout.to_sql(name='dropout', con=conn)
```

```
data_dropout.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110 entries, 0 to 109
Data columns (total 14 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   State_UT             110 non-null    object
 1   year                 110 non-null    object
 2   Primary_Boys         110 non-null    object
 3   Primary_Girls        110 non-null    object
 4   Primary_Total        110 non-null    object
 5   Upper Primary_Boys   110 non-null    object
 6   Upper Primary_Girls  110 non-null    object
 7   Upper Primary_Total  110 non-null    object
 8   Secondary _Boys      110 non-null    object
 9   Secondary _Girls     110 non-null    object
 10  Secondary _Total     110 non-null    object
 11  HrSecondary_Boys     110 non-null    object
 12  HrSecondary_Girls    110 non-null    object
 13  HrSecondary_Total    110 non-null    object
dtypes: object(14)
memory usage: 12.2+ KB
```

```python
# SQL command
sql_command = """WITH query1 AS
                (
                    SELECT State_UT AS state, SUM(Primary_Total) AS Primary_Rate, SUM("Upper Primary_Total") AS Upper_Primary_Rate, SUM("Secondary _Total") AS Secondary_Rate , SUM(HrSecondary_Total) AS HrSecondary_Rate
                    FROM dropout
                    GROUP BY State_UT
                )
                SELECT * , (Primary_Rate + Upper_Primary_Rate + Secondary_Rate + HrSecondary_Rate) AS Total
                FROM query1
                GROUP BY state
                ORDER BY Total DESC
                """

# execute the statement
cur.execute(sql_command)

# store all the fetched data in the ans variable
query_result = cur.fetchall()

# convert the results to a pandas DataFrame
df = pd.DataFrame (query_result,columns=['State','Primary_Rate' , 'Upper_Primary_Rate','Secondary_Rate', 'HrSecondary_Rate', 'Total_Rate'])

# Print the first five rows of the DataFrame
df
```

	State	Primary_Rate	Upper_Primary_Rate	Secondary_Rate	HrSecondary_Rate	Total_Rate
0	Nagaland	32.09	35.61	79.85	39.45	187.00
1	Assam	29.04	24.76	84.26	11.93	149.99
2	Odisha	9.36	10.42	128.90	0.00	148.68
3	Mizoram	47.17	30.08	62.00	6.91	146.16
4	Karnataka	7.31	11.41	93.67	17.29	129.68
5	Daman & Diu	1.96	6.67	63.65	51.60	123.88
6	Meghalaya	29.94	21.21	71.30	0.00	122.45
7	Tripura	7.08	7.79	79.01	26.91	120.79

HYPOTHESIS#2

BOYS HAVE BETTER CHANCE OF BETTER QUALITY OF EDUCATION THAN GIRLS

Group 9 .Sep 2020

```
[6]: # Create database
     dbname = 'IndianEducationDb'
     conn = sqlite3.connect(dbname + '.sqlite')

     cur = conn.cursor()

     # Load the data to SQL Database
     enroll.to_sql(name='Enroll', con=conn)
```

```
[7]: #enroll.State_UT = enroll.State_UT.str.capitalize()
     enroll = enroll.replace('NR', np.nan, regex=True)
     enroll = enroll.replace('@', np.nan, regex=True)
```

```
enroll.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110 entries, 0 to 109
Data columns (total 14 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   State_UT            110 non-null    object
 1   Year                110 non-null    object
 2   Primary_Boys        110 non-null    float64
 3   Primary_Girls       110 non-null    float64
 4   Primary_Total       110 non-null    float64
 5   Upper_Primary_Boys  110 non-null    float64
 6   Upper_Primary_Girls 110 non-null    float64
 7   Upper_Primary_Total 110 non-null    float64
 8   Secondary_Boys      110 non-null    float64
 9   Secondary_Girls     110 non-null    float64
 10  Secondary_Total     110 non-null    float64
 11  Higher_Secondary_Boys 108 non-null  object
```

Girls Enrollment

```python
sql_command = """WITH query1 AS
                (
                    SELECT State_UT AS state, SUM(Primary_Girls) AS GirlsPrimary_Rate, SUM("Upper_Primary_Girls") AS GirlsUpper_Primary_Rate, SUM("Secondary_Girls") AS GirlsSecondary_Rate , SUM(Higher_Secondary_Girls) AS
                    FROM Enroll
                    GROUP BY State
                )
                SELECT * , (GirlsPrimary_Rate + GirlsUpper_Primary_Rate + GirlsSecondary_Rate + GirlsHrSecondary_Rate) AS TotalGirls
                FROM query1
                GROUP BY State
                ORDER BY TotalGirls DESC
                """

# execute the statement
cur.execute(sql_command)

# store all the fetched data in the ans variable
query_result = cur.fetchall()

# convert the results to a pandas DataFrame
df = pd.DataFrame (query_result,columns=['State','GirlsPrimary_Rate' , 'GirlsUpper_Primary_Rate','GirlsSecondary_Rate', 'GirlsHrSecondary_Rate', 'TotalGirls'])

# Print the first five rows of the DataFrame
df
```

	State	GirlsPrimary_Rate	GirlsUpper_Primary_Rate	GirlsSecondary_Rate	GirlsHrSecondary_Rate	TotalGirls
0	Delhi	341.49	404.67	320.16	280.13	1346.45
1	Sikkim	326.95	420.50	351.30	223.14	1321.89
2	Mizoram	365.33	377.60	327.31	181.99	1252.23
3	Manipur	422.09	370.35	266.08	186.17	1244.69
4	Himachal Pradesh	301.44	311.55	335.35	292.02	1240.36
5	Arunachal Pradesh	381.10	373.48	262.57	196.42	1213.57
6	Meghalaya	421.70	397.30	259.65	113.65	1192.30
7	Tripura	332.80	367.61	359.14	118.43	1177.98

Boys Enrollment


```python
sql_command = """WITH query2 AS
                (
                    SELECT State_UT AS state, SUM(Primary_Boys) AS BoysPrimary_Rate, SUM("Upper_Primary_Boys") AS BoysUpper_Primary_Rate, SUM("Secondary_Boys") AS BoysSecondary_Rate , SUM(Higher_Secondary_Boys) AS BoysHr
                    FROM Enroll
                    GROUP BY State
                )
                SELECT * , (BoysPrimary_Rate + BoysUpper_Primary_Rate + BoysSecondary_Rate + BoysHrSecondary_Rate) AS TotalBoys
                FROM query2
                GROUP BY State
                ORDER BY TotalBoys DESC
                """

# execute the statement
cur.execute(sql_command)

# store all the fetched data in the ans variable
query_result = cur.fetchall()

# convert the results to a pandas DataFrame
df = pd.DataFrame (query_result,columns=['State','BoysPrimary_Rate' , 'BoysUpper_Primary_Rate','BoysSecondary_Rate', 'BoysHrSecondary_Rate', 'TotalBoys'])

# Print the first five rows of the DataFrame
df
```

	State	BoysPrimary_Rate	BoysUpper_Primary_Rate	BoysSecondary_Rate	BoysHrSecondary_Rate	TotalBoys
0	Mizoram	377.59	382.91	320.40	173.48	1254.38
1	Delhi	326.20	354.70	306.82	261.64	1249.36
2	Sikkim	352.37	410.64	307.84	176.14	1246.99
3	Himachal Pradesh	296.29	307.19	350.25	292.42	1246.15
4	Manipur	407.06	354.01	269.85	204.91	1235.83
5	Arunachal Pradesh	385.51	359.89	273.14	195.18	1213.72
6	Tripura	329.78	357.75	353.12	136.51	1177.16
7	Lakshadweep	240.08	289.26	342.88	294.83	1167.05

HYPOTHESIS #3

THE QUALITY OF EDUCATION IN LESS DEVELOPED STATES THAN THE QUALITY OF
WELL DEVELOPED STATES

```
display(data_dropout.head)
```

```
<bound method NDFrame.head of          State_UT     year Primary_Boys Primary_Girls Primary_Total  \
0       A & N Islands  2012-13         0.83          0.51          0.68
1       A & N Islands  2013-14         1.35          1.06          1.21
2       A & N Islands  2014-15         0.47          0.55          0.51
3      Andhra Pradesh  2012-13          3.3          3.05          3.18
4      Andhra Pradesh  2013-14         4.31          4.39          4.35
..                ...      ...          ...           ...           ...
105       West Bengal  2013-14         3.44          2.37          2.91
106       West Bengal  2014-15         2.13          0.79          1.47
107         All India  2012-13         4.68          4.66          4.67
108         All India  2013-14         4.53          4.14          4.34
109         All India  2014-15         4.36          3.88          4.13

     Upper Primary_Boys Upper Primary_Girls Upper Primary_Total  \
0          Uppe_r_Primary                1.09                1.23
1                     NR                1.54                0.51
2                   1.44                1.95                1.69
3                   3.21                3.51                3.36
4                   3.46                4.12                3.78
..                   ...                 ...                 ...
105                 5.63                 3.1                4.31
106                 5.84                2.88                 4.3
107                  2.3                4.01                3.13
108                 3.09                4.49                3.77
109                 3.49                 4.6                4.03

     Secondary _Boys Secondary _Girls Secondary _Total HrSecondary_Boys  \
0               5.57             5.55             5.56            17.66
1               8.36             5.98              7.2            18.94
2              11.47             8.16             9.87            21.05
3              12.21            13.25            12.72             2.66
4              11.95            13.37            12.65            12.65
..               ...              ...              ...              ...
105            16.73            19.77            18.34             8.03
106            16.33            19.06             17.8             8.18
107            14.54            14.54            14.54               NR
108            17.93            17.79            17.86             1.48
109            17.21            16.88            17.06             0.25
```

```
In [70]:   1   #Dropout Rates
           2   DropoutRate = """SELECT State_UT,year, HrSecondary_Total from SDA.SchoolsDropout where year
           3   cursor.execute(DropoutRate)
           4
           5   myresult = cursor.fetchall()
           6
           7   # convert the results to a pandas DataFrame
           8   df = pd.DataFrame (myresult)
           9
          10   # Print the first five rows of the DataFrame
          11   df
          12   #for x in myresult:
          13   #     print(x)
```

	State_UT	year	HrSecondary_Total
0	Daman & Diu	2014-15	40.48
1	Arunachal Pradesh	2014-15	18.42
2	Delhi	2014-15	17.32
3	A & N Islands	2014-15	16.93
4	Goa	2014-15	13.91
5	Jammu & Kashmir	2014-15	12.65
6	Sikkim	2014-15	11.76
7	Chandigarh	2014-15	10.55
8	Dadra & Nagar Haveli	2014-15	9.47
9	Tripura	2014-15	8.93
10	West Bengal	2014-15	8.11
11	Himachal Pradesh	2014-15	7.41
12	Gujarat	2014-15	7.04

```python
GenderCom = """ SELECT DISTINCT SDA.data_girls_toilet.State_UT, SDA.data_girls_toilet.year
 SDA.data_girls_toilet.All_Schools AS Girls_ALL, SDA.data_boys_toilet.All_Schools AS Boys_A
FROM SDA.data_girls_toilet left JOIN SDA.data_boys_toilet ON data_girls_toilet.year=SDA.dat
WHERE SDA.data_girls_toilet.State_UT = 'All India' AND SDA.data_boys_toilet.State_UT = 'All
"""

cursor.execute(GenderCom)


myresult = cursor.fetchall()
# convert the results to a pandas DataFrame
df = pd.DataFrame (myresult)


# Print the first five rows of the DataFrame
df
```

	State_UT	year	Girls_ALL	Boys_ALL
0	All India	2013-14	91.23	86.56
1	All India	2014-15	93.08	88.62
2	All India	2015-16	97.52	97.02

```python
In [76]:   1  # SQL command
           2  sql_command = """SELECT distinct SDA.SchoolsWithComp.State_UT, SDA.SchoolsWithComp.year,  S
           3                  """
           4
           5  # execute the statement
           6  cursor.execute(sql_command)
           7
           8  # store all the fetched data in the ans variable
           9  query_result = cursor.fetchall()
          10
          11  # convert the results to a pandas DataFrame
          12  df = pd.DataFrame (query_result)
          13
          14  # Print the first five rows of the DataFrame
          15  df
```

	State_UT	year	Schools with computers	Schools with Elect
0	Jharkhand	2014-15	9.71	17.98
1	Assam	2014-15	9.83	22.40
2	Bihar	2014-15	8.19	25.22
3	Jammu And Kashmir	2014-15	17.56	26.39
4	Meghalaya	2014-15	10.64	26.72
5	Madhya Pradesh	2014-15	14.58	28.29
6	Tripura	2014-15	15.11	28.58
7	Odisha	2014-15	13.72	29.73
8	Manipur	2014-15	25.23	30.66
9	Arunachal Pradesh	2014-15	24.68	38.60
10	Nagaland	2014-15	35.68	40.94

MAP MIND FOR OUE

PREDICTION

Statistics of Indian School Education

The objective goal is to burrow a few realities about the nature of Indian Education from 2013 - 2016 and help individuals of India and Government of India to see better about the Indian Schools to take quick activities if necessary.

The quality of education in less developed states than the quality of well developed states

boys have better chance of better quality of education than girls

Indian School Education

Most of the schools that lack electricity are in poor states. However, there are some school in states with high GDP that doesn't have electricity, we believe that those schools are rural areas.

Is the opposite than our thoughts families enroll girls more than boys.
- one of the reasons because female born more than boys
- The Other reason is it might because in the poor states boys supports their family by working and provide for their families.

student in less developed states drop out more to support their families

After analyzing the data we found a relation between the rate of dropout and the rate of poverty. Thus, we believe that students in poor states dropout in order to support their families.